图书情报工作
第65卷 第8期 2021年4月

全球 OA 科技期刊 APC 监测与异常预警模型研究*

■ 芮啸[1] 赵展一[1,2] 王昉[1] 陈雪飞[1] 黄金霞[1,2]

[1] 中国科学院文献情报中心 北京 100190 [2] 中国科学院大学经济管理学院 北京 100190

摘 要：[目的/意义]及时掌握全球 OA 科技期刊 APC 发展态势与异常情况，有利于避免国内科研经费流失，辅助决策，为国内确定面向开放获取和开放科学的发展路径提供参考。[方法/过程]使用最小二乘法拟合 OA 期刊的 APC 单价和影响力指数之间的函数关系，并设置 95% 的置信区间以识别溢价期刊、非溢价期刊，基于此原理建立全球 OA 科技期刊 APC 监测与异常预警模型。[结果/结论]该模型包括自动采集、监测分析、异常预警模块，以中国科学院某物理学领域研究所发文的 OA 期刊数据代入模型计算，发现溢价期刊占比 43.5%，有 4.46% 的异常 APC 需要预警，但 90% 的论文发表在了性价比较高的非溢价期刊上，并且非溢价期刊均不在 Beall's List 名单中，溢价期刊分类符合 GoOA 和 DOAJ 的收录情况，验证了模型的有效性和可靠性。

关键词：OA 期刊 论文处理费用 溢价期刊 异常预警模型 开放科学
分类号：G255
DOI：10.13266/j.issn.0252-3116.2021.08.005

随着开放获取（Open Access，OA）运动愈演愈烈，全球开放出版市场规模也不断壮大。据测算，2019 年中国论文总量约 195 万篇[1]，中国约占全球论文总量 18%（数据来源：Web of Science；计算方式：61 万/336 万≈18%）[2]，由此得出 2019 年全球论文总量约 1 000 万篇。开放获取期刊目录（Directory of Open Access Journals，DOAJ）平台显示 2019 年全球 OA 论文总量约 60 万篇[3]，其中金色 OA 论文约占全球 OA 论文总量 30%（数据来源：Web of Science；计算方式：35 万/120 万≈30%），由此得出 2019 年全球 OA 论文总量约 200 万篇，占全球论文总量的 20%。

2019 年至今，全球达成多项大型 OA 转换协议，如 Elsevier 与卡内基梅隆大学、Taylor Francis 与挪威、Wiley 与瑞士 Bibsam 联盟，以及 2020 年 1 月初，Springer Nature 与德国学术联盟 PROJECT DEAL 签订了世界上金额最大的 OA 转换协议[4]等。截至 2020 年 2 月 1 日，开放获取期刊目录（Directory of Open Access Journals，DOAJ）已收录 14 200 多种 OA 期刊，461 多万篇 OA 论文；Web of Science（WoS）显示 5 670 多万篇学术论文中大约 18%（960 多万篇）是 OA 论文，包括 377 万多篇金色 OA 论文。以金色 OA 论文为例，全球范围内，以金色 OA 方式出版的文章比例由 2012 年的 11% 增长至 2016 年的 19%，英国作者所撰写的文章以金色 OA 方式出版的比例从 2012 年的 11% 增长至 2016 年的 30%，这一时期的年均增长率超过 28%；如果算上复合 OA 期刊，2016 年，英国一半以上的文章发表在 OA 期刊上[5]。可见 OA 期刊的发展已成不可阻挡的趋势。开放出版有利地促进了学术研究成果的发布、传播和利用，但大多数出版机构需要收取论文处理费用（Article Process Charge，APC）来维持商业模式正常运行，伴随着 OA 出版的迅猛发展，APC 的增长也突飞猛进，有研究指出，APC 总量的增长速度是刚开始收取 APC 时的三倍[6]，部分营利性出版商更是以此为契机制定激进的营销策略，引发了一系列问题。一方面，掠夺性出版商以大肆圈钱为目的，使不经严格审稿的论文进入学术生态系统，如印度出版商 OMICS 不对稿件进行任何评审及修改就发布，未经科学家们允许就将其姓名列在编辑委员会中，还擅自在其官方网站中将 PubMed Central 和 Medline 列为索引数据库，美国联邦贸易委员会已于 2019 年 3 月对其判处了 5 000 万美元

* 本文系国家社会科学基金项目“全球 OA 科技期刊出版大数据监测模型研究”（项目编号：18BTQ059）研究成果之一。

作者简介：芮啸（ORCID：0000-0003-4042-4310），馆员，硕士；赵展一（ORCID：0000-0001-8116-9139），博士研究生；王昉（ORCID：0000-0002-1069-1541），副研究馆员，硕士；陈雪飞（ORCID：0000-0003-4945-3695），馆员，硕士；黄金霞（ORCID：0000-0001-8705-0067），研究馆员，博士，通讯作者，E-mail：huangjx@mail.las.ac.cn。

收稿日期：2020-11-09 修回日期：2021-02-07 本文起止页码：42-50 本文责任编辑：杜杏叶

的罚款[7];另一方面,开放出版的无节制扩张导致了严重的科研经费流失现象,有学者选取 8 种代表性国际 OA 期刊作为统计来源,发现中国作者仅在 2017 年就对其 APC 贡献了将近 4 亿元,平均贡献率高达 50% 以上,这些 OA 期刊简直是为中国作者"量身定制"的[8]。此外,有研究表明完全 OA 期刊 APC 的平均价格在 2 000美元以下,而混合型文章在研究型大学的研究人员的出版物上平均约为 3 000 美元[9],其中的期刊 APC 价格合理性几乎无法判断,出版商在价格制定中行使自由裁量权。

国际方面意识到成本和价格是影响 OA 转型的关键因素,已启动一些 APC 监测专项以了解 APC 与 OA 出版市场的动态情况,如 2015 年德国启动"Open APC"项目,通过获取成员机构 APC 单价、总额等数据促进 OA 出版费用的透明化,改善学术出版市场功能失调情况,增强学术出版的可持续性[10],2013 年英国联合信息系统委员会(Joint Information System Committee,JISC)实施"出版总成本"(Total Cost of Ownership,TCO)项目,委托 Information Power Ltd 公司收集英国高等教育机构关于期刊订阅和 APC 支出的相关数据[11],2014 年 JISC 推出"Monitor Local"服务集中记录 APC 支出和 OA 资助政策,从成本和政策的角度高效跟踪科研机构的 OA 发展情况[12],另外 DOAJ 等第三方平台也会收集和提供 OA 期刊的 APC 信息,但上述项目或平台大多起到延时记录和汇总报告的作用,缺乏异常预警功能,因此,本文通过探究 APC 异常条件设计了全球 OA 科技期刊 APC 监测与异常预警模型,并以中国科学院某研究所为例验证了模型的有效性和可靠性。

2020 年 2 月中国科技部发布《关于破除科技评价中"唯论文"不良导向的若干措施(试行)》,对论文发表费用支出、审核进行了严格规定:对于单篇论文发表支出超过 2 万元人民币的,需经该论文通讯作者或第一作者所在单位学术委员会对论文发表的必要性审核通过后,方可在国家科技计划项目专项资金中列支。对于发表在"黑名单"和预警名单学术期刊上的论文,相关的论文发表支出不得在国家科技计划项目专项资金中列支[13]。在 OA 期刊出版的未来发展中,资助者将扮演重要的作用。希望本文的研究结果能为资助者的战略制定有所贡献。

1 研究目标

国际期刊出版界通常用"Publication fees"或"Article Process Charges(APC)"表示论文从加工处理到最终发表的费用总和,后者通常用于 OA 期刊[14],支付方式主要包括 4 种:①作者单篇支付或会员支付;②机构会员全额支付或与作者分担支付;③联盟成员分担支付;④其他资助模式。随着 OA 期刊运营模式愈加成熟,出版商的 APC 定价策略也多种多样,影响因素包括 OA 期刊的类型、所属学科、质量、影响力、所属国家/地区、技术成本以及出版机构的业务规模、营销策略[15]等。

有研究表明 OA 期刊的影响力与期刊声誉、APC 存在相关性,如戴琦等依据 B. Klein & K. B. Leffler 提出的声誉激励机制建立了以影响力为核心的 OA 期刊声誉风险评估机制,具体包括影响因子、被引半衰期、h5 指数、特征因子分值等指标[16]。B. C. Björk 全面调查了 OA 期刊 APC 和论文质量(以影响力和论文数量为测度)之间的关系,作者选用不同学科引文的 SNIP 值进行了标准化处理,发现其和 APC 的相关系数是 0.4[17]。刘烜贞等通过调查 266 种生命科学领域 SCI 期刊的影响因子和 APC 关系,发现二者基本呈正相关关系[18]。另外,笔者所在团队从 2013 开始跟踪全球 OA 期刊发展动态并尝试做了一些评价、监测工作,初步发现 OA 期刊的影响力是 APC 的重要影响因素[19]。据此,本文假设 OA 期刊的 APC 与影响力具有正相关关系,使用影响力指标构建单因子模型。

进一步,邀请国内大型科技期刊编辑部主任,中科院研究所科研人员代表,中科院战略研究院研究员,北京大学图书馆研究员,科协中国科技期刊卓越行动计划成员等共 5 名专家,针对 OA 期刊评价各级指标(共 70 个)进行打分,每项 0 – 5 分,单项最高 25 分。此处只摘选专家对 OA 期刊影响力指标的打分分值。根据专家打分结果(见图 1),并考虑数据可获取和量化程度,本文选取谷歌 h5 指数和 SCI 影响因子作为 OA 期刊影响力的衡量指标,二者分数相同,所占权重也相同。

根据影响力水平把 OA 期刊分为高影响期刊和低影响期刊,根据 APC 收费水平把 OA 期刊分为高价期刊和低价期刊。所有 OA 期刊可被组合为高价高影响力期刊(A1)、高价低影响力期刊(A2)、低价低影响力期刊(A3)和低价高影响力期刊(A4)。

用函数 $y = f(x)$ 表示 OA 期刊的 APC 和期刊影响力之间的关系,通过平移坐标轴,函数穿过坐标轴交叉点,在不考虑置信区间的情况下,将 $y = f(x)$ 函数图像上方的期刊定义为溢价期刊,溢价期刊溢出的 APC 部

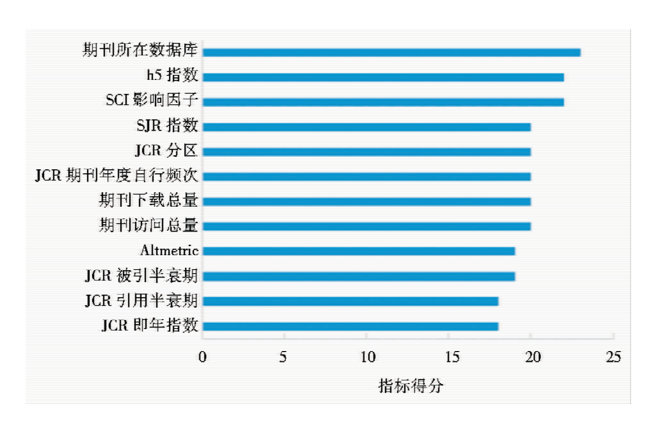


图 1 影响力指标专家打分结果

分为异常 APC,下方的期刊为非溢价期刊。溢价期刊中一定包含所有 A2 和部分 A1、A3,非溢价期刊中一定包含所有 A4 和部分 A1、A3,见图 2。

基于上述原理,本文建立的全球 OA 科技期刊 APC 监测与异常预警模型拟实现 3 个目标:①识别出溢价期刊、非溢价期刊并对溢价期刊进行预警;②监测溢价期刊、非溢价期刊的发文量情况;③监测 APC 总量并预警异常 APC 总量。

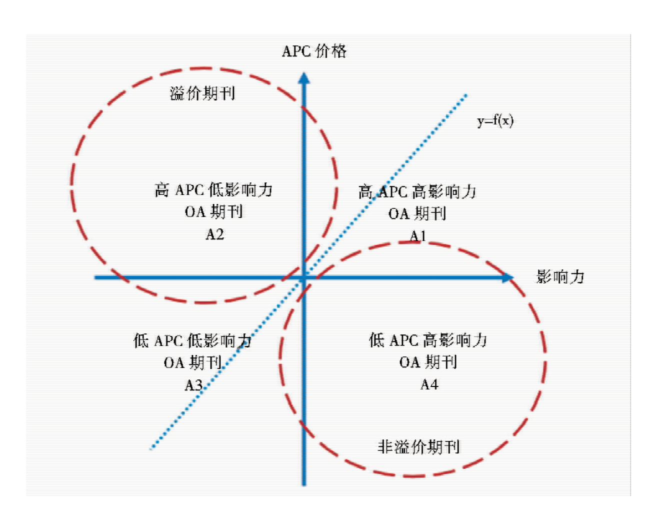


图 2 A1、A2、A3、A4 与溢价、非溢价期刊的关系分布

2 模型设计

全球 OA 科技期刊 APC 监测与异常预警模型面向所有学科、领域设计,包括自动采集模块,监测分析模块和异常预警模块,数据配置和流动过程如表 1 所示:

表 1 全球 OA 科技期刊 APC 监测与异常预警模型功能模块

自动采集模块	采集数据	信息来源
	APC、发文量 影响因子 h5 指数 所有数据直接存入数据库	期刊官网 科睿唯安期刊引证报告(Journal Citation Reports,JCR)[20] 谷歌学术[21]
监测分析模块	模块细分	输出结果
	数据清洗和融合 函数拟合 期刊分类	期刊 APC 和影响力 期刊 APC 和影响力的拟合函数 y = f(x) 溢价期刊和非溢价期刊
异常预警模块	预警类别	预警内容
	溢价期刊 异常 APC 其他异常变化	性价比低,不推荐投稿 异常 APC 总量 APC 总量、异常 APC 总量、期刊总量、期刊发文总量等

在自动采集模块,根据所需数据的类型、来源编写程序,自动采集期刊的发文量、APC 单价、影响因子、h5 指数等数据,直接存入数据库以待监测分析和异常预警模块调用。各指标数据具体情况如表 2 所示:

表 2 自动采集模块所需数据信息

数据名称	数据类型	数据来源
APC 单价	数值型	期刊官网或出版社官网,一般位于"开放获取""作者须知""投稿须知"等栏目中
发文量	数值型	如果以期刊为单位,则期刊发文总量在期刊官网中;如果以研究机构为单位,则以该机构在该期刊上发文量为准。
影响因子	数值型	科睿唯安期刊引证报告
h5 指数	数值型	谷歌学术

在监测分析模块,使用自动采集到的数据计算出 OA 期刊的 APC 单价和影响力指标,运用最小二乘法拟合出 APC 单价和影响力指标的一元线性回归函数,根据函数图像划分溢价期刊和非溢价期刊。各部分计算方式如下:

(1)为避免不同量级的影响,分别对影响因子和 h5 指数进行归一化,使数据通过函数变换映射到 0 ~ 1 之间得到标准化处理。并且,数据落在特定区间内能够使寻优过程更加平缓,增强模型的普适性。归一化公式为:

$$影响力指标 = Q_1(I_m/I_{max}) + Q_2(H_m/H_{max}) \qquad 公式(1)$$

其中:

Q_1,Q_2 分别表示影响因子和 h5 指数对影响力指数所占的权重,$Q_1 = Q_2$ 且 $Q_1 + Q_2 = 1$;

I_m 表示期刊影响因子的数值;

I_{max} 表示所采集期刊中影响因子的最大值；

H_m 表示期刊 h5 指数的数值；

H_{max} 表示所采集期刊中 h5 指数的最大值；

I_{max} 和 H_{max} 是指所选领域、机构 OA 期刊对应的影响因子、h5 指数最大值，具体数值会随样本范围的变化而变化。

(2)由于本文假设影响力和 APC 呈正相关关系，所以尝试采用最小二乘法拟合 OA 期刊的 APC 和影响力之间的一元线性回归关系，使误差平方和达到最小[22]。令 OA 期刊影响力指标为自变量 x，APC 单价为因变量 y，x、y 之间的拟合函数为：

$$y = ax + b \qquad \text{公式(2)}$$

其中：

a 表示自变量 x 的系数，其计算公式为：

$$a = \frac{\sum_{i=1}^{n} y_i(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad \text{公式(3)}$$

b 表示常数项，其计算公式为：

$$b = \bar{y} - a\bar{x} \qquad \text{公式(4)}$$

$\bar{x}$ 和 $\bar{y}$ 分别表示 x 和 y 的平均值，其计算公式为：

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}, \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \qquad \text{公式(5)}$$

需要注意的是，OA 期刊的收费策略十分多样，不同数据库、不同期刊、不同论文，甚至是同一篇论文也可能因选择不同的开放条件而使 APC 有所不同[10]。因此，若 OA 期刊或出版商未直接给出篇均 APC 单价，本文将按照字数(5 000 字)、页数(10 页)、图表数(5 个)自动预估 APC 总额；若 APC 定价存在地区差异，本文以亚洲地区为准；若单价单位不统一，本文参照 2019 年 12 月 20 日汇率转换标准统一转换为美元进行计算。不考虑其他优惠、作者顺序等因素。

在异常预警模块，首先对上一模块中的溢价期刊进行预警，其次根据 APC 单价和发文量计算出溢价期刊的异常 APC 总量，计算方式为：

$$\text{异常 APC 总量} = \sum_{i=1}^{m}[(A_i - A_i')\ N_i] \qquad \text{公式(6)}$$

其中：

m 表示溢价期刊总量；

N 表示期刊发文总量；

A_i 表示溢价期刊的 APC 单价；

A_i' 表示溢价期刊的计算 APC，其代表将该溢价期刊的影响力 x_i 带入 $y = f(x)$ 得到的 y，也即计算 APC。

最后，可通过对历年数据进行函数拟合预测 APC 总量和异常 APC 总量的未来发展趋势。

3 模型验证

3.1 APC 单价和影响力指数拟合函数

国内中国科学院某物理领域研究所对 OA 期刊的支持度较高，本文以 2000－2019 年中国科学院某研究所在 OA 期刊上发表的论文为例，分析其 APC 异常情况。从 WoS 中导出该所科研人员曾发过 OA 论文的 377 条期刊数据，去除非 OA 期刊、会议、专著、停刊期刊，共得到 246 个 OA 期刊。由于 APC 为 0 的期刊，仅对于作者而言是 0，出版社层面是因为该期刊 APC 由基金或机构支付，并且具体支付金额无从测算，所以将 APC 为 0 的期刊直接归为非溢价期刊并不带入拟合总体进行拟合计算。去除后共有 218 个 OA 期刊。

由于无法得到期刊历年的 APC 数据，只能通过 APC 单价的年均增长率计算往年的 APC。Bo-Christer Björk 等通过研究表明 APC 单价的年均增幅为大约 5%[16]，本文以该数值倒推出 2019 年之前的各年度 APC 总量，计算方法如下：

$$APC_n = APC_{2019} * 0.95^{2019-n} \qquad \text{公式(7)}$$

其中：

n 表示年份；

APC_n 表示某期刊 n 年的 APC 单价；

APC_{2019} 表示某期刊 2019 年的 APC 单价；

通过最小二乘法拟合出中科院某研究所发文期刊所对应 APC 单价和影响力的一元回归方程为 y = 1 505.5x + 1 745.3，函数图像见图 3。

此外也尝试使用多元回归方程进行拟合，如图 4、图 5 分别为二元和三元回归的效果图。

二元函数拟合结果的表达式为 y ＝ －33.863x2 ＋ 1 554.1x ＋ 1 733.3，其函数图像几乎与一元函数拟合结果重合，三元函数拟合结果的表达式为 y ＝ 1 305.4x3 － 3 126.7x2 ＋ 3 486.1x ＋ 1 436.5，其函数图像与一元函数拟合结果仍然不大，所以本文最终选择一元函数拟合并进行后续计算。

运用 T 检验计算 P-value = $1.74E^{-13}$，在一元线性回归中，F 值与 T 值都与相关系数 R 等价，因此 Significance F、P-value 都可以用来表示函数拟合效果，且 P-value 一般要小于 0.05，越小效果越显著，当 P-value 小于 0.01 时为非常显著。本例中 P-value 远远小于 0.01，表明拟合效果非常显著，即 APC 与影响力呈非常显著的正相关关系。该式中，常数项 1 745.3 说明当影响力为 0 时，正常的 APC 单价应该为 1 745.3，低于 1 745.3 的期刊不会出现溢价；一次项系数为 1 505.5，

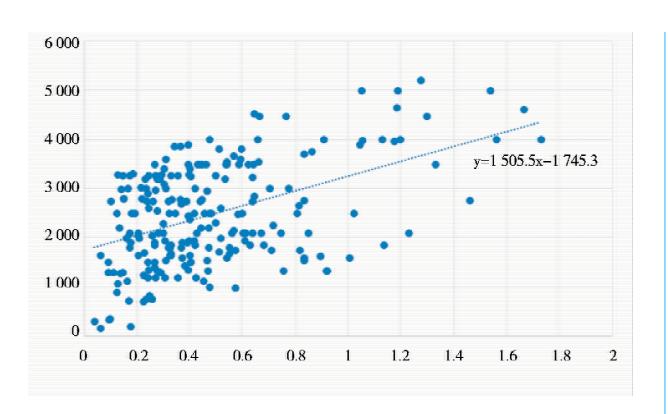


图 3　APC 单价和影响力指数拟合函数

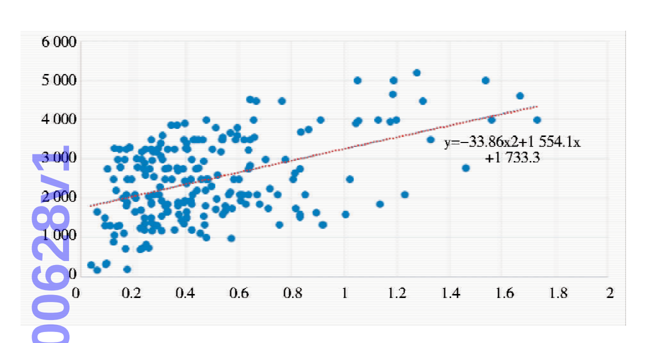


图 4　APC 单价和影响力指数二元拟合函数

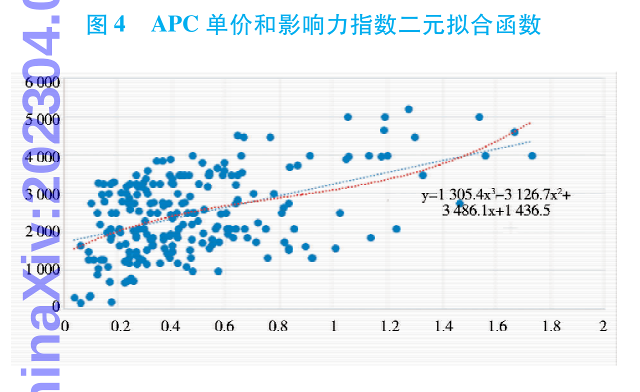


图 5　APC 单价和影响力指数三元拟合函数

说明影响力每增加 0.1，APC 可上涨 150.55，而影响力的最大值为 2，说明 APC 的标准上限应该为 4 756.3，当期刊收费高于 4 756.3 时，为溢价期刊。

从图 6 中可以看出，很多溢价期刊在函数图像附近，甚至 APC 仅仅高出几美元的期刊也被归为溢价期刊，欠缺合理性，因此本文引入 95% 置信区间用以增强函数的说服力和灵活性，即落在 95% 置信区间内的期刊不归为溢价期刊。图 6 中浅蓝色区域 为 95% 置信区间，区间内的点不被计入 APC 异常范围内。

3.2　溢价期刊的识别与第三方验证

获取到期刊的 APC 单价、h5 指数、影响因子、发文量数据后，计算得出截至 2019 年该所发文的异常 APC 总量为 239759 美元，有 107 个 OA 期刊需要预警，详见表 3：①溢价期刊有 107 个，占比 43.5%，非溢价期刊有 139 个，占比 56.5%，溢价与非溢价期刊数量上占比接近 1∶1；②溢价期刊发文 390 篇，占比 9.72%，非溢价期刊发文 3 621 篇，占比 90.28%，溢价与非溢价期刊在发文数量上占比接近 1∶9；③溢价期刊发文花费 APC 约 123.6 万美元，占比 23.03%，非溢价期刊发文花费 APC 约 413.2 万美元，占比 76.97%，溢价与非溢价期刊在发文所花费 APC 数量上占比接近 1∶4，需要预警的异常 APC 占比 4.46%。

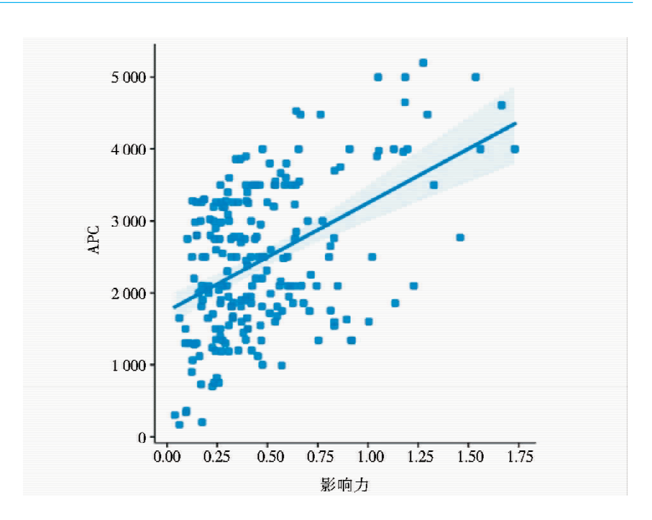


图 6　落入 95% 置信区间的非溢价期刊

表 3　中科院某研究所发文的 OA 期刊 APC 异常预警情况

期刊类型	期刊数/种	发文量/篇	APC 总金额/美元	异常 APC 金额/美元
溢价期刊	107	390	1 236 248	239 759
非溢价期刊	139	3 621	4 131 972	0
全部期刊	246	4 011	5 368 220	239 759

结合①②可发现，溢价与非溢价期刊在数量上接近，但发文数量悬殊，说明该所大多数论文发表在了非溢价期刊上，投稿者拥有较好的溢价期刊辨识能力；结合②③可发现，非溢价期刊发文数量远高于溢价期刊，但 APC 支出占比的差距不大，说明溢价期刊的平均 APC 远高于非溢价期刊。

GoOA[23] 和 DOAJ[2] 是国内外具备 OA 期刊严格质量控制机制的平台，“Beall’s List”（比尔黑名单）列出了大量质量低下的 OA 期刊[24]，曾得到学术界广泛认可（已于 2017 年停止更新）。将模型对期刊的分类结果与上述第三方数据进行对比（见表 4），以验证模型的有效性，发现：①GoOA 收录溢价期刊 6 个，占溢价期刊总数的 5.61%，收录非溢价期刊 42 个，占非溢价期刊总数的 30.22%；②DOAJ 收录溢价期刊 7 个，占溢价期刊总数的 6.54%，收录非溢价期刊 63 个，占非溢价期刊总数的 45.32%；③Beall’s List 收录溢价期刊 4 个，且没有收录任何非溢价期刊。由此可以发现，模型

对优劣期刊,尤其是 Beall's List 或 DOAJ 可疑期刊名单中的期刊具有良好的鉴别效果,可靠性强。

表 4　中科院某研究所发文 OA 期刊被第三方收录情况

期刊类型	期刊数	GoOA 收录	DOAJ 收录	黑名单收录
	数量/种	数量/种	数量/种	数量/种
溢价期刊	107	6	7	4
非溢价期刊	139	42	63	0
全部期刊	246	48	70	4

3.3　监测溢价期刊、非溢价期刊的发文量和 APC 单价、总量情况

通过将该所近 6 年所有 OA 期刊的发文与 APC 支出数据代入模型,得出非溢价期刊的 APC 总量、溢价期刊的异常 APC 总量(见表 4),以及中科院某研究所近 6 年溢价期刊、非溢价期刊的分布情况(见图 7)。

2014－2019 年间,该所发文的非溢价期刊数量占比一直保持在 70% 左右,非溢价期刊发总量占比一直保持在 90% 左右;非溢价期刊大的 APC 占总比一直都保持在 77% 左右。该所绝大多数的文章发表和 APC 的投入都在非溢价期刊上且能够保持稳定,说明该所科研人员发文的 OA 期刊质量比较受信赖。另外,该所只投过 4 本黑名单期刊,占比 1.63%,发文量占比 0.35%,并且 2017 年后就没有在这 4 本期刊发过论文,非溢价期刊比例也从 65% 上升到 72% 左右,说明该所科研人员对性价比高的期刊辨识能力在 2017 年后不断提升。

表 5　中科院某研究所近 6 年 OA 期刊发文与 APC 支出状况

年份/年	非溢价期刊			溢价期刊				总计		
	期刊数/种	发文量/篇	APC 总额/元	期刊数/篇	发文量/篇	APC 总额/元	异常 APC 总量/元	期刊数/篇	发文量/篇	APC 总额/元
2014	37	306	286 553	16	28	71 629	21 393	53	334	358 182
2015	34	369	360 527	15	35	87 973	22 170	49	404	448 500
2016	44	430	365 999	23	38	107 845	28 394	67	468	473 844
2017	38	386	363 223	15	36	46 623	28 823	53	422	409 846
2018	49	526	617 990	19	44	125 709	36 413	68	570	782 841
2019	37	496	683 917	15	61	222 992	64 742	52	557	906 909

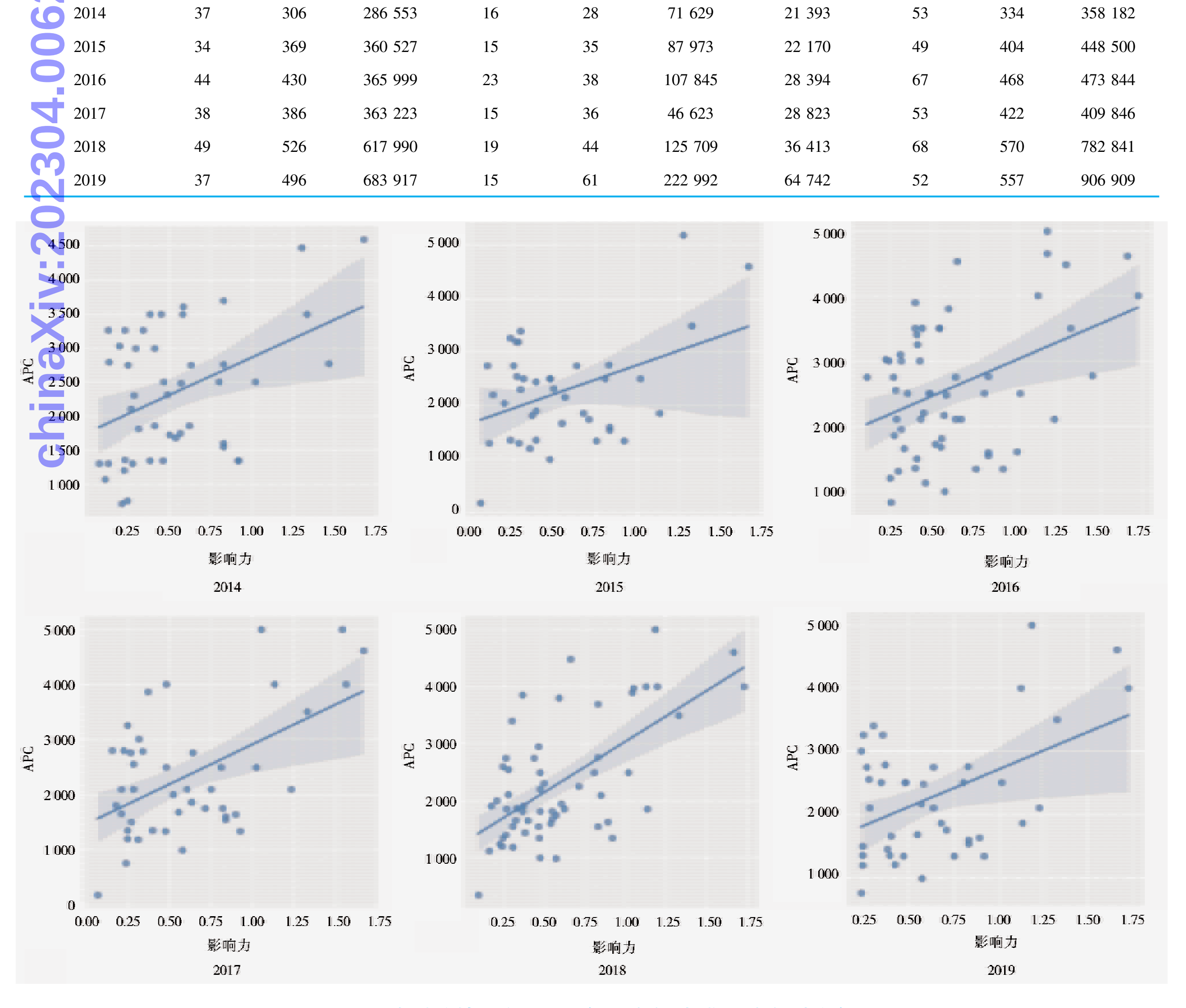


图 7　中科院某研究所近 6 年溢价期刊、非溢价期刊分布

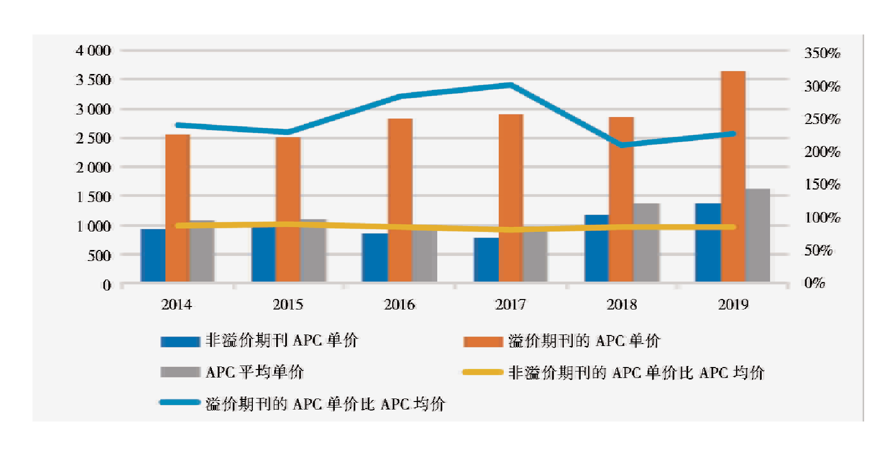


图8 各类期刊 APC 单价均价、占比年度变化

非溢价期刊、溢价期刊和所有期刊的平均 APC 单价呈上升趋势,但是溢价期刊和非溢价期刊的 APC 单价对 APC 均价的比值却不同,非溢价期刊的 APC 单价稳定在 85% 左右,溢价期刊的 APC 单价在 200% – 300% 的比值内波动,平均比值为 255%。也就是说,溢价期刊的 APC 单价一般是非溢价期刊的 APC 单价的 3 倍,足以可见溢价期刊的收费水平之高。

3.4 监测并预警异常 APC 总量

近 6 年,异常 APC 在溢价期刊 APC 和总 APC 中的占比分别稳定在 28%、6% 左右;溢价期刊发文量占比也一直稳定在 9% 左右;溢价期刊总数在 2016 年出现一次高峰,占比达到 34%,2017 年回落并保持在 29% 左右;溢价期刊 APC 占总比从 2015 开始上升后在 2017 年大幅下降并在 2018 年回升。可以看出,虽然溢价期刊数量和 APC 总量的占比一直在变化,但是从溢价期刊中计算出的异常 APC 相对于 APC 总量能够一直保持稳定。说明该模型能够有效过滤不稳定因素的影响,从而掌握异常 APC 的内在规律。

总体来看,中国科学院某研究所 APC 支出总量和发文总量总体呈上升态势,并在近两年增幅加快,接下来的几年在 OA 期刊上的花费和发文量可能会进入快速增长的状态中;非溢价期刊的发文量、APC 单价、APC 总量都保持在相对稳定的状态,并且发文量和期刊总量都远大于溢价期刊,呈现出良好的发展趋势;从定义来看,溢价期刊就是高价低影响力的期刊,数据本身也显示出溢价期刊的 APC 是非溢价期刊的 3 倍左右,且不稳定,另外第三方平台或机构也很少收录溢价期刊。异常 APC 虽由溢价期刊产生,但在 APC 总量的占比一直很稳定,只有非溢价期刊 APC 占比大量减少或溢价期刊 APC 占比大量增多才会导致异常 APC 占 APC 总量的比值上升。所以机构可将该比值作为异常 APC 的警戒值,且可推广到更大范围的监测中去,一旦异常 APC 的比值高于该值,监测方就向决策方发出预警,并通过数据溯源找到异常原因。

4 结语

本文使用最小二乘法拟合 OA 期刊的 APC 单价和影响力指数之间的函数关系,并设置 95% 的置信区间以识别溢价期刊、非溢价期刊,基于此原理设计了包含自动采集、监测分析、异常预警模块的全球 OA 科技期刊 APC 监测与异常预警模型。该模型识别出中科院某研究所的科研人员在 2000 – 2019 年间发文的溢价期刊占比 43.5%,发文量占比 9.72%,有 4.46% 的异常 APC 需要预警,并且所有非溢价期刊均被 GoOA、DOAJ 白名单收录;2014 – 2019 年间的数据显示,该所科研人员将 90% 的论文发表在了性价比较高的非溢价期刊上,并且异常 APC 相对于 APC 总量的占比一直保持 6% 左右,可将 6% 作为预警阈值。综合来看,单因子模型效果良好,既实现了模型预期目标,也印证了前人对影响力和 APC 相关关系的研究。

该模型旨在监测全球各学科、各机构 OA 科技期刊 APC 异常情况并预警,但本文以中科院某研究所作为试验样本,若条件允许,并且基于笔者团队 GoOA 工作掌握的全球 OA 科技期刊评价数据,未来可继续扩大样本范围,验证模型对一个学科、一个国家乃至全球 OA 学术科技期刊 APC 的监测和预警效果,甄别不良定价策略,有效避免国家或机构科研经费流失,为国家建立 APC 定价规则,OA 期刊发展体系以及向开放科

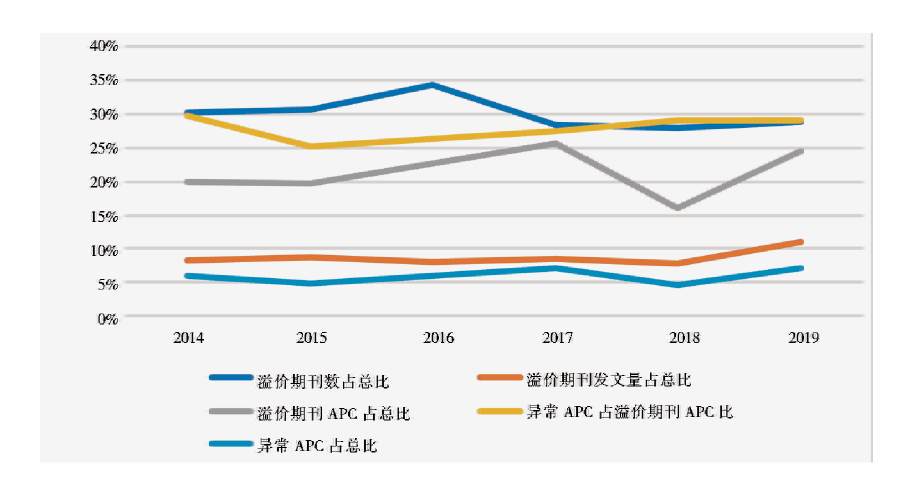


图 9　中科院某研究所近 6 年溢价期刊及发文、异常 APC 占比年度变化

学生态系统过渡提供参考。另外,该模型除了可以对 OA 期刊集合的 APC 情况进行监测预警外,也可分辨单刊状态。只要将新增单刊信息放入模型中,更新拟合函数和图像,便可实现。该模型还有更多应用场景有待讨论。

致谢:本文在进行模型和数学公式的设计中得到张国涵博士的指导,在此表示感谢!最初向张博士请教时,他是在中国科学院数学与系统管理研究院开展博士后工作,后来他到北京邮电大学理学院任教,仍然对我们的工作提出了很多宝贵的建议。

参考文献:

[1] 2020 中国统计年鉴[EB/OL]. [2021－01－10]. http://www.stats.gov.cn/tjsj/ndsj/2020/indexch.htm.

[2] Web of science[EB/OL]. [2021－01－10]. http://isiknowledge.com/wos.

[3] Dirctory of open access journals[EB/OL]. [2021－01－10]. https://doaj.org.

[4] Springer nature and germany's projekt DEAL finalise world's largest transformative open access agreement[EB/OL]. [2020－01－09]. https://www.webwire.com/ViewPressRel.asp? aId = 253040.

[5] 初景利,闫雪. 数字出版与高校图书馆的应对——对 Springer Nature 集团和剑桥大学图书馆的调查[J]. 图书馆论坛,2020,40(9):173－179.

[6] 王岚. 付费 OA 出版与 OA 出版开支管理研究[J]. 山东图书馆学刊,2016(05):95－100,106.

[7] FTC hits predatory scientific publisher with a ＄50 million fine [EB/OL]. [2019－04－09]. https://arstechnica.com/science/2019/04/ftc-hits-predatory-scientific-publisher-with-a-50-million-fine/.

[8] 江晓原,穆蕴秋. "开放存取运动":科学出版乌托邦的背后——Nature 实证研究之六[J]. 上海交通大学学报(哲学社会科学版),2018,26(3):5－20.

[9] SOLOMON D, BJORK B C. Article processing charges for open access publication - the situation for research intensive universities in the USA and Canada[J]. Science editor and publisher, 2018, 2(2/4):89－106.

[10] PIEPER D, BROSCHINSKI C. OpenAPC: a contribution to a transparent and reproducible monitoring of fee-based open access publishing across institutions and nations[J]. Insights, 2018, 31: 39.

[11] 张丽英. 英国出版总成本项目对中国开放出版的启示[J]. 情报资料工作,2019,40(3):98－103.

[12] Monitor local:easily record open access compliance and cost[EB/OL]. [2020－10－05]. https://monitor.jisc.ac.uk/local/about/.

[13] 科技部印发《关于破除科技评价中"唯论文"不良导向的若干措施(试行)》的通知[EB/OL]. [2020－02－23]. http://www.most.gov.cn/xxgk/xinxifenlei/fdzdgknr/fgzc/gfxwj/gfxwj2020/202002/t20200223_151781.html.

[14] 余敏. 欧美出版社开放存取期刊论文处理费研究[J]. 出版科学,2016,24(5):106－110.

[15] 么媛媛,郑建程. 开放出版的 APC 模式研究[J]. 中国科技期刊研究,2014,25(8):1016－1020.

[16] 戴琦,袁曦临. 开放获取期刊的学术声誉风险及其预警研究[J]. 中国科技期刊研究,2018,29(11):1063－1071.

[17] BJORK B C,SOLOMON D. Article processing charges in OAjournals: relationship between price and quality[J]. Scientometrics, 2015,103(2):373－385.

[18] 刘烜贞,陈静. 开放获取期刊出版费及其对学术交流的影响[J]. 中国科技期刊研究,2015,26(12):1244－1249.

[19] 武虹,陈雪飞,王昉,等. 全球 OA 期刊出版大数据监测方法[J]. 现代情报,2019,39(2):145－151.

[20] Journal citation reports[EB/OL]. [2020－12－01]. https://clarivate.com.tw/products/journal-citation-reports.

[21] Google scholar[EB/OL]. [2021 - 01 - 01]. https://scholar.google.com.
[22] 邹乐强. 最小二乘法原理及其简单应用[J]. 科技信息, 2010(23):282 - 283.
[23] 开放获取论文一站式发现平台[EB/OL]. [2021 - 01 - 01]. http://gooa.las.ac.cn/external/index.jsp.
[24] Jeffrey beall. criteria for determining predatory open-access publishers[EB/OL]. [2019 - 12 - 21]. https://web.archive.org/web/20161130184313/https://scholarlyoa.files.wordpress.com/2015/01/criteria-2015.pdf.

作者贡献说明：
芮啸：完成数据的采集和清洗，方法的调研和选取，模型的设计和实验，论文的主要部分撰写；
赵展一：完成前期调研，数据处理，模型修改，论文统稿撰写；
王昉：提供研究思路，研究方法指导，提供修改意见；
陈雪飞：指导模型设计，提供修改意见；
黄金霞：提供总体研究框架，撰写论文大纲，模型设计思路，提供修改意见。

Research on APC Monitoring and Abnormal Warning Model of Global OA Sci-Tech Journals

Rui Xiao[1] Zhao Zhanyi[1,2] Wang Fang[1] Chen Xuefei[1] Huang Jinxia[1,2]
[1] National Science Library, Chinese Academy of Sciences, Beijing 100190
[2] School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190

Abstract: [**Purpose/significance**] Grasping the development and abnormal situation of global OA Sci-Tech journals' APC in time is helpful to avoid the loss of domestic research funding, assist decision-making, and provide a reference for China to determine the development path for open access and open science. [**Method/process**] This paper used the least square method to fit the functional relationship between the APC unit price and influence index of OA journals, and set a 95% confidence interval to identify premium journals and non-premium journals. Based on this principle, a global OA Sci-Tech journals' APC monitoring and anomaly early warning model was established. [**Result/conclusion**] The model includes automatic acquisition, monitoring analysis, and anomaly early warning modules. Take OA journals data published by an institute of physics in the Chinese Academy of Sciences as an example, it was found that premium journals accounted for 43.5%, and 4.46% of abnormal APCs required early warning. However, 90% of the papers were published in non-premium journals with high cost performance, and non-premium journals were not in the Beall's List, premium journal classification conforms to the inclusion of GoOA and DOAJ, which verified the validity and reliability of the model.

Keywords: Open Access journals article processing charges premium journals abnormal early warning model open science